DOCUMENT RESUME

ED 430 802 SE 062 587

AUTHOR Wang, Jianjun; Staver, John
TITLE An Examination of Science Items in TIMSS.
PUB DATE 1999-03-00
NOTE 10p.; Paper presented at the Annual Meeting of the National Association for Research in Science Teaching (72nd, Boston, MA, March 28-31, 1999).
PUB TYPE Reports - Research (143) -- Speeches/Meeting Papers (150)
EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Achievement Rating; Elementary Secondary Education; Foreign Countries; *Item Analysis; *Science Tests; Test Construction; *Test Items
IDENTIFIERS *Third International Mathematics and Science Study

ABSTRACT

Development of the test instrument in the Third International Mathematics and Science Study (TIMSS) was based on the expertise of many researchers, including "distinguished scholars from 10 countries" who participated on the TIMSS Subject Matter Advisory Committee. However, a close examination of the TIMSS Science items suggests that not all the items measure student science achievement. Thus, cautions are raised from this study to urge researchers and policymakers to take a more mindful stance against potential misinterpretation of TIMSS scores in international comparisons. (Author/WRM)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

SCOPE OF INTEREST NOTICE
The ERIC Facility has assigned this document for processing to: SE
In our judgment, this document is also of interest to the Clearinghouses noted to the right. Indexing should reflect their special points of view. TM

ED 430 802

Running head: TIMSS Item Examination

An Examination of Science Items in TIMSS

Jianjun Wang
Department of Teacher Education
School of Education
California State University
9001 Stockdale Highway
Bakersfield, CA 93311-1099

John Staver
Center for Science Education
College of Education
Kansas State University
Manhattan, KS 66056

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY
J. Wang
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
1

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
This document has been reproduced as received from the person or organization originating it.
Minor changes have been made to improve reproduction quality.
Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

Paper presented at the 1999 annual meeting of the National Association of Research in Science Teaching (NARST), March 28-31, Boston, MA.

BEST COPY AVAILABLE

SE025587

An Examination of Science Items in TIMSS

Abstract

Development of test instrument in the Third International Mathematics and Science Study (TIMSS) was based on expertise of many researchers, including "distinguished scholars from 10 countries" in its Subject Matter Advisory Committee (SMAC). However, a close examination of the TIMSS science items suggests that not all the items measure student science achievement. Thus, cautions are raised from this study to urge researchers and policymakers taking more mindful stance against potential misinterpretation of TIMSS scores in the international comparison.

An Examination of Science Items in TIMSS

The Third International Mathematics and Science Study (TIMSS) is a collaborative research project sponsored by the International Association for Evaluation of Educational Achievement (IEA). The existing debate on TIMSS findings was mainly focused on various confounding variables, such as variations among the international curricula, influences of cultural traditions, and differences in school enrollments (Stedman, 1997a, b; Bracey, 1997). In contrast, few researchers have closely examined the academic contents of the TIMSS instrument. In part, this was because the academic tests were endorsed by the TIMSS Subject Matter Advisory Committee (SMAC), "including distinguished scholars from 10 countries" (Beaton, Mullis, Martin, Gonzalez, Kelly, & Smith, 1996, p. A-5). As a result, Peak (1996) noted that "TIMSS is a fair and accurate comparison of mathematics and science achievement in participating nations" (p. 14). On the other hand, Gerald W. Bracey (1998), an outspoken researcher against the TIMSS horse race, also admitted that "The published sample items strike me as quite reasonable things to teach students" (p. 686).

To date, two thirds of the TIMSS items were disseminated on a web page (http://wwwcsteep.bc.edu/timss1/database.html). Discussions of the item problems must be confined among those items released to the public. To maintain academic integrity and honesty, some of the released TIMSS items have been cross-examined in this article to underscore five potential problems within the TIMSS measurement. Since the TIMSS results have drawn dramatic attention in the United States (http://whitehouse.dm.net/library/1998/10836.TXT), the item examination may inform policymakers and the American public to take a more mindful stance on some of the TIMSS outcomes.

Efforts on the instrument development were documented in a TIMSS Technical Report (Martin & Kelly, 1996). In the past, international comparisons were based on results from multiple choice tests (Rotberg, 1995). In TIMSS, about one-third of students' response time was devoted to free-response questions. Despite the format improvement, the TIMSS instrument illustrated the following problems in the areas of science and mathematics:

. Not all free-response scores reflect student science achievement

According to the TIMSS Technical Report (Martin & Kelly, 1996), specific rubrics were developed for each free-response item using a two-digit coding system. The grading rubrics, however, contained scientific problems. For instance, a TIMSS question reads,

> The water level in a small aquarium reaches up to a mark A. After a large ice cube is dropped into the water, the cube floats and the water level raises to a new mark B. What will happen to the water level as the ice melts? Explain your reasoning. (Item # G11, http://wwwcsteep.bc.edu/timss1/Items.html)

This item was included in the 12th grade physics test. Without information about the experimental temperature, the melting process could take a number of hours, which might introduce significant impact of evaporation along with the liquidation process. Ignoring the effect of evaporation, students could reach an answer of "same level". On the other hand, considering the missing matter through evaporation, students could develop an answer of "lower level". To a certain extent, the second answer may come from more considerate students. According to the TIMSS rubrics, the second answer was one of the incorrect responses, and thus, earned zero credit!

. Not all the mistakes were isolated

It is understandable that isolated mistakes may happen in a large scale study like TIMSS.

However, the TIMSS instrument indicated that similar mistakes repeated across different surveys. For instance, the aforementioned problem was re-appeared in the TIMSS instrument at other grade levels. In its science test for the 7th and 8th graders, an item reads,

> A glass of water with ice cubes in it has a mass of 300 grams. What will the mass be immediately after the ice has melted? Explain your answer. (Lie et al., 1996, p. 7-11)

This item was even listed in the TIMSS Technical Report as an exemplary item (Martin & Kelly, 1996). Still, the coding guide only award credit to an answer of 300 grams, and thus, punished those students who have thoughtfully considered the missing matter through evaporation during the melting process (Lie et al., 1996, p. 7-11). Unfortunately, this kind of problems was not discussed in the TIMSS quality control document (Mullis & Smith, 1996)).

. Not all science items have only one correct choice

If the above example illustrated that the TIMSS grading missed a correct answer, some TIMSS items listed more than one correct choice. The following item was quoted from the TIMSS science test at the 3rd and 4th grade levels:

> Seeds develop from which part of a plant?
> A. Flower B. Leaf C. Root D. Stem
> (Item # P9, http://wwwcsteep.bc.edu/timss1/Items.html)

The content of seed and plant was covered in many widely-used elementary science textbooks (e.g., Mallinson, Mallinson, Smallwood, & Valentino, 1984). A total of two groups of seed plants were introduced at the basic level. Specifically, according to Mallinson, Mallinson, Smallwood, and Valentino (1984), "One of these groups is made up of seed plants that have cones. ... The second group of seed plants is made up of seed plants that have flowers" (p. 29). Thus, even at the beginning stage of a science course, students knew that choice A,

Flower, was not the only answer. Cones, as another kind of seeds, were developed from the stem part of the plant. Therefore, good students had to struggle between choices A and D to match a single answer in the rubrics! In this circumstance, the bad question may have indeed confused a good student.

. Not all TIMSS scores are grounded on student levels of cognitive development

In the TIMSS science test, one question at the 3rd and 4th grade levels states:

> The Sun is bigger than the Moon, but they appear to be about the same size when you look at them from the Earth. Why is this?
> (Item # Y1, http://wwwcsteep.bc.edu/timss1/Items.html)

The answer hinges on a notion that the Sun is farther away than the Moon. However, according to the TIMSS code, if a student "Refers to the sun being higher up than the moon", he or she will receive a zero score. To describe the distance difference in the sky, the two words, higher and farther, could be used interchangeably by most 3rd and 4th graders. This characteristic is typically for students at the concrete operation level of cognitive development (Piaget, Chomsky, & Piatelli-Palmarini, 1980). Unfortunately, the TIMSS grading system failed to consider the level of student cognitive development.

. Not all TIMSS items are developed through proper collaborations between mathematics and science educators

TIMSS is the only IEA project which assessed mathematics and science achievement concurrently. TIMSS items also covered mathematics applications in the field of science. However, not all the items are free of misconceptions. The problem is quite obvious in a mathematics item at the 7th and 8th grades. The item reads,

> A chemist mixes 3.75 milliliters of solution A with 5.625 milliliters of solution B to form a new solution. How many milliliters does this new solution contain?

(Item # K2, http://wwwcsteep.bc.edu/timss1/Items.html)

The answer from the TIMSS code was 9.375. Apparently, the item writer assumes that when the two solutions are mixed, the volume will be additive. The additive assumption, nonetheless, is not true in many cases, particularly when the answer carries many significant figures after the decimal point. For instance, if one solution was mainly made by alcohol and the other by water, then the volume should be less than 9.375! Thus, this item is grounded on assumption which could be undermined by counter-examples in science.

In summary, at a cost to U.S. taxpayers of $51 million, TIMSS findings have drawn great public interest. While many researchers participated in discussions of TIMSS results, few questioned the TIMSS' measure of science and mathematics achievement. With all due respect to the distinguished scholars of the TIMSS Subject Matter Advisory Committee, the content examination indicated that not all TIMSS items truly reflected student achievement in science. Since the TIMSS data have been widely disseminated in the public, researchers and policymakers should analyze TIMSS items carefully, and avoid potential misinterpretation of TIMSS scores in international comparisons.

References

Baker, D. P. (1997). Surviving TIMSS. Phi Delta Kappan, 79 (4), 295-300.

Beaton, A., Mullis, I., Martin, M., Gonzalez, E., Kelly, D. & Smith, T. (1996). Mathematics achievement in the middle school years. Chestnut Hill, MA: Boston College.

Bracey, G. (1997). On comparing the incomparable: A response to Baker and Stedman. Educational Researcher, 26 (3), 19-26.

Bracey, G. (1998). TIMSS, rhymes with 'dims,' as in 'witted'. Phi Delta Kappan, 79 (9), 686-687.

Lie, S., Taylor, A. & Harmon, A. (1996). Scoring techniques and criteria. In M. Martin & D. Kelly (Eds.), Third international mathematics and science study: Technical report. Chestnut Hill, MA: Boston College.

Mallinson, G., Mallinson, J., Smallwood, W., & Valentino, C. (1984). Science. San Carlos, CA: Silver Burdett.

Martin, M., & Kelly, D. (1996). Third international mathematics and science study: Technical report. Chestnut Hill, MA: Boston College.

Mullis, I., & Smith, T. (1996). Quality control steps for free-response scoring. In M. Martin & I. Mullis (Eds.), Third international mathematics and science study: Quality assurance in data collection. Chestnut Hill, MA: Boston College.

Peak, L. (1996). Pursuing excellence: A study of U.S. eighth-grade mathematics and science teaching, learning, curriculum, and achievement in international context (NCES No. 97-198). Washington, DC: U.S. Government Printing Office.

Peak, L. (1997). Pursuing excellence: A study of U.S. fourth-grade mathematics and

science achievement in international context (NCES No. 97-255). Washington, DC: U.S. Government Printing Office.

Piaget, J., Chomsky, N., & Piatelli-Palmarini, M. (1980). Language and learning : the debate between Jean Piaget and Noam Chomsky. Cambridge, MA: Harvard University Press,

Rotberg, I. (1995). Myths about test score comparisons. Science, 270, 1446-1448.

Stedman, L. (1997a). International achievement differences: An assessment of a new perspective. Educational Researcher, 26 (3), 4-15.

Stedman, L. (1997b). Deep achievement problems: The case for reform still stands. Educational Researcher, 26 (3), 27-29.

Takahira, S., Gonzales, P., Frase, M., & Salganik, L. H. (1997). Pursuing excellence: A study of U.S. twelfth-grade mathematics and science achievement in international context (NCES No. 98-049). Washington, DC: U.S. Government Printing Office.

Wang, J. (in press). International achievement comparison: Interesting debates on inconclusive findings. School Science and Mathematics.